Big Data Research 27 (2022) 100292

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Big Data Analytics and Visualization in Traffic Monitoring

Chiara Bachechi, Laura Po *, Federica Rollo

"Enzo Ferrari" Engineering Department, University of Modena and Reggio Emilia, Italy

ARTICLE INFO

Article history:
Received 15 March 2021
Received in revised form 10 August 2021
Accepted 28 October 2021
Available online 10 November 2021

Keywords:
Traffic
Time series
Air quality maps
Real-time data
Spatio-temporal data
Smart cities

ABSTRACT

This paper presents a system that employs information visualization techniques to analyze urban traffic data and the impact of traffic emissions on urban air quality. Effective visualizations allow citizens and public authorities to identify trends, detect congested road sections at specific times, and perform monitoring and maintenance of traffic sensors. Since road transport is a major source of air pollution, also the impact of traffic on air quality has emerged as a new issue that traffic visualizations should address. Trafair Traffic Dashboard exploits traffic sensor data and traffic flow simulations to create an interactive layout focused on investigating the evolution of traffic in the urban area over time and space. The dashboard is the last step of a complex data framework that starts from the ingestion of traffic sensor observations, anomaly detection, traffic modeling, and also air quality impact analysis. We present the results of applying our proposed framework on two cities (Modena, in Italy, and Santiago de Compostela, in Spain) demonstrating the potential of the dashboard in identifying trends, seasonal events, abnormal behaviors, and understanding how urban vehicle fleet affects air quality. We believe that the framework provides a powerful environment that may guide the public decision-makers through effective analysis of traffic trends devoted to reducing traffic issues and mitigating the polluting effect of transportation.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

"Transportation is a complex world. It is a mix of technologies, social behaviors, choices of single users and stochastic events, nested within a geographical, environmental and economic scenario." [1] With the rapid development of transportation systems, traffic has become an important part of human life and significantly influenced the quality of life since an estimated average of 40% of the population spends at least 1 hour on the road every day [2]. The European Environment Agency estimates that road transport contributes to excessive concentrations of about 70% of nitrogen dioxide (NO_2) [3].

For a smart sustainable city, big data visualization implies turning a large amount of urban data into useful knowledge for enhanced decision-making and deep insights concerning various urban domains, such as transport, mobility, traffic, environment, energy, land use, waste management, education, healthcare, public safety, and governance [4]. To achieve sustainable development goals defined by the 2030 Agenda [5], traffic analytics need to be coupled with the air quality impact to address a change towards more sustainable mobility.

This paper aims to tackle the problem of traffic monitoring and traffic analysis in space and time and to unveil the effects on urban air quality according to the mobility choices made. In this paper, we proposed a novel traffic data analytics platform that comprises tools for real-time traffic monitoring, effective traffic analytics over time, and means for understanding how changes in the urban vehicle fleet can mitigate the impact of traffic on air pollution. Trafair Traffic Dashboard (TTD) is a visual analytics tool to allow the exploration of real-time and historical data and envision the polluting impact of traffic through a unified interactive interface. The dashboard is backed by several data analysis techniques, such as sensor observation visualization, anomaly detection, traffic simulation analysis, allowing the exploration of behavioral similarities between sensors or neighborhoods, the visual detection of unusual events, and the simulation of traffic flow on a new hypothetical vehicle fleet scenario defined by the user. The applicability of the proposed dashboard in facilitating decision-making is demonstrated in two case scenarios, using real traffic data sets. The TTD has been conceived within the Trafair project, which aims to help public administrations and citizens to understand traffic flow in order to improve air quality in six European cities. A traffic dashboard was needed to visualize, analyze and understand traffic data and the impact of road traffic on urban air quality. The data visualized are harvested from the city sensor network or generated by two simulation models: a traffic model and an air pollution dispersion model. Pursing the idea of relying only on Open Source

* Corresponding author.
E-mail addresses: chiara.bachechi@unimore.it (C. Bachechi), laura.po@unimore.it (L. Po), federica.rollo@unimore.it (F. Rollo).

https://doi.org/10.1016/j.bdr.2021.100292
2214-5796/© 2021 Elsevier Inc. All rights reserved.

software, all the software employed in the framework is free and open source.

The rest of the paper is structured as follows. Section 2 reviews related work on time series, geospatial data, spatio-temporal data, and traffic-related data management and visualizations. Section 3 outlines the project goals and lists the Trafair Traffic Dashboard requirements. The framework and the data flow that feeds the dashboard are described in Section 4. Section 5 describes the technological choices for the implementation of the dashboard and details all the views that fulfill the requirements. In Section 6, we present use cases that demonstrate the advanced capabilities of our visual analytics dashboard focused on urban traffic monitoring on two real scenarios: the city of Modena (Italy), and Santiago de Compostela (Spain). Limitations are discussed in Section 7. Section 8 sketches conclusion and future work.

2. Related work

Time series and geospatial data visualization have attracted enormous attention in the last decade. A time series is a chronological sequence of observations on a particular variable. Time-series data have particular features; they are large in data size, have high dimensionality, and update continuously. Various mining tasks can be performed on time series data: clustering, anomaly detection, visual representation, short-term and long-term prediction, and others. Time series analysis is a tricky task [6] that handles issues such as representation and indexing, similarity measure, segmentation, pattern discovery, and classification.

Geospatial (or spatial georeferenced) data describe features on the Earth's surface. Geospatial data are managed through a Geographic Information System (GIS), a conceptualized tool that allows the user to create interactive queries to store, edit and visualize spatial data, often jointly with non-spatial data. Specific features of geographical data preclude the use of general-purpose data mining algorithms. Therefore, for several decades an active area of research is Spatial Data Mining (SDM). SDM [7] consists of extracting knowledge, spatial relationships, and, in particular, identifying previously unknown patterns.

Data that have both spatial and temporal dimensions are called Spatio-Temporal data (ST), this is also the case of traffic data. ST data have auto-correlation in both space and time: instances are related to each other and their properties vary in different spatial regions and time periods. Managing ST data means handling issues related to both temporal and spatial data. Moreover, the visualization choice matter, as it affects how the users perceive similarity in ST [8].

The traffic-related data, covered in this article, are point reference data, raster data, or spatial maps with a temporal dimension. Following the approach of [9], our solution was to treat spatial locations as objects (the sensors or the roads in the network) and using the measurements collected and produced by the model over time to define their features.

Literature provides some guidelines for selecting proper exploratory techniques for ST data depending on the characteristics of the data and the analysis goals. In [10], authors consider spatial (where), temporal (when), and object (what) components of ST data individually and in combination with each other, for example determining characteristics of spatial objects at a given moment or analyzing the behavior of an object or location by observing it on a series of maps referring to consecutive time moments.

The negative impact of urban traffic on air pollution is well known. In [11], the effect of NO_x emissions of Euro 6 Diesel cars on the NO_2 pollution in 8 European cities has been addressed. Another approach [12] attempted in visualizing changes in air pollution levels resulting from changes in traffic conditions due to the implementation of various urban transport schemes.

2.1. Traffic dashboards

Several visualization platforms have been developed to provide insights from traffic-related data. In Table 1, we provide a comparison of the traffic dashboards analyzed in the following.

TA-Dash [13] is part of the Data4UrbanMobility project [18] that provide an event-based framework of mobility information. TA-Dash allows non-expert users to perform analysis of congestion patterns, in particular, to identify the impact of planned special events and to detect structured spatial-temporal dependencies in an urban road network. The utility of the dashboard has been demonstrated in the city of Hannover, Germany.

Extensive surveys on traffic management key requirements have been employed in the Traffic Management as a Service (TMaaS) project [14],[1] which aims to provide interactive traffic management dashboards for the traffic operators along with personalized mobility services for the citizens. TMaas uses an architecture designed to be flexible, lightweight, data-independent, and vendor-independent. It has been implemented for the city of Ghent, Belgium, and will become operational in other 4 cities in the coming months.

The dashboard used in the city of Porto [15] has a strong focus on providing the delay and speed profiles for a specific road at different periods of the day. Moreover, the pollutant concentration trend is provided and compared with the traffic trend of each road. It has been conceived as a tool for city managers for improving urban traffic, and is part of the Sensing and Serving a Moving City project.[2]

Some other dashboards are focus on specific aspects, like for example, the UK Road traffic accident (RTA) platform [16]) that provide an in-depth analysis of the traffic accidents. The interactive visual analytics (IVA) platform [17], demonstrated in an area of San Jose, California, allows the exploration of real-time and historical data and the prediction of future traffic-related attributes (such as volume and average vehicle speed) given the current traffic state thanks to advanced data analysis algorithms.

TTD mainly focuses on providing average trends and weekly, monthly, and yearly statistics for traffic flow and speed. TTD did not focus on traffic forecasting because it was out of the scope of the Trafair project. TTD allows exploring both data coming from traffic sensors and data obtained through simulation processes for urban traffic and air pollution. Since the pollutant levels do not depend only on the emissions generated by traffic and are heavily affected by the weather condition and winds, we avoid a direct comparison between the pollutants level and the traffic profile in a street as in [15].

3. Project description and requirements

Trafair Traffic Dashboard was developed within the scope of the Trafair project, a European project that aims to develop innovative and sustainable services combining air quality, weather conditions, and traffic flows data to produce new information for the benefit of citizens and government decision-makers.

Trafair raises awareness among public administrations about the air quality within an urban environment and the pollution caused by traffic. The project aims at monitoring air quality by using sensors in 6 cities [19–21] and making air quality predictions thanks to two simulation models: the traffic model and the air pollutant dispersion model. Traffic data concern both data coming from sensor and data generated by a traffic simulation model. The dashboard empowers traffic managers of public administrations to

[1] https://drive.tmaas.eu/nl/.

[2] https://www.cmuportugal.org/s2movingcity/.

Table 1
Traffic dashboards comparison.

		TA-Dash [13]	TMaaS [14]	Porto Traffic dashboard [15]	UK RTA platform [16]	IVA platform [17]	TTD
Data employed	historical traffic information	x				x	x
	open geospatial data		x	x	x	x	x
	traffic events		x		x	x	
	city events	x					
	government or commercial data		x	x			x
Traffic view	interactive geographic maps	x	x	x	x	x	x
	real-time traffic monitoring		x			x	x
	trajectory or pattern oriented view	x					
	road section traffic profile			x			
	day of the week statistics			x			x
	historical data visualizations					x	x
	average traffic flow evolution hour by hour					x	x
	average annual daily traffic maps					x	x
	road traffic forecast					x	
Additional features	traffic accidents				x	x	
	city event traffic impact	x					
	personalized alerts for citizens		x				
	decision support for city manager		x	x	x		x
	traffic-environment correlation			x			x
	anomaly detection					x	x

understand urban traffic flow in space and time, vehicle emissions, and the consequent degradation of ambient air quality. Insights can leverage for more sustainable urban planning decisions.

The TTD requirements, listed in the following, have been collected during the Trafair annual meeting held in Zaragoza in November 2019; they are the result of the needs expressed by public administrations and the dialogue and discussion with environmental experts, mobility managers, and researchers who have worked together to investigate urban traffic and its impact on air quality:

R0 automatically updating visualization of real-time traffic flow in the city,
R1 visualization of sensor positions, last measurements, sensor's behavior, and the reliability of the measurements;
R2 statistical information about traffic sensor measurements;
R3 visualization of the average day of the week trends and other aggregations in space and time;
R4 historical trends of traffic flow in the urban area;
R5 traffic impact on urban air quality.

Since smart cities are ingesting more and more data and the urban sensor network is set to increase, the data model and the designed architecture should ensure scalability. Scalability means the ability to add or remove sensor devices without affecting the system's availability. The proposed solution is designed to be applied in different cities that may have different data sources and sizes and may need to model the data structure differently. For this reason, a flexible structure is a powerful lever for increasing adaptability in different contexts.

4. Data flow and Trafair framework

Data that are displayed in TTD are involved in a complex system of data pre-processing, ingestion, cleaning, and modeling. This system consists of several steps, starting from the road network ingestion and sensor data acquisition, passing through the traffic modeling, going by an intermediate level till the visualization of the traffic flow and statistics in the dashboard.

Fig. 1 provides a visualization of the data flow and the architecture of the Trafair framework. As displayed on the left side of the figure, the road network of the area of interest together with the position and measurements of traffic sensors are stored in a PostgreSQL database that will be detailed in Section 4.1. A data cleaning process (Section 4.2) removes unreliable sensor observations thanks to the speed-flow correlation filter and the anomaly detection algorithm. Cleaned data are taken as input by the traffic model (Section 4.3) to simulate traffic flows in each road of the area of interest. Also, the model output is stored in the database. The simulated traffic flows generate vehicular emissions that are taken as input by the air pollutant dispersion model (Section 4.4). The output of this model is uploaded on GeoServer as GeoTIFFs (georeferenced rasters). The data cleaning process as well as the traffic model and the air pollutant dispersion model are time and resource consuming processes; therefore, they are executed on High Performance Computing (HPC) platform. Materialized views that aggregate and summarize the sensor measurements and model outputs are created on the PostgreSQL database. The views are exposed through several GeoServer layers, as will be described in Section 4.5. These layers are queried by the TTD, implemented with Angular, to show interactive maps and graphs (Section 5).

4.1. Data modeling

The data platform consists of a PostgreSQL database that exploits two extensions: PostGIS and Timescale. PostGIS[3] allows handling geospatial data by including specific data types (e.g. geometry, geography, and raster) and by storing points, lines, and polygons. Timescale[4] is a time-series extension for providing fast analytics, and scalability on PostgreSQL DB. This extension can be applied on each table with a temporal reference, i.e. a column with a date/time. When applied, the table is transformed into a hypertable and its content is split into several chunks based on the value of the timestamp and the time interval chosen by the database administrator.

Timescale automatically creates a b-tree index on the column containing the timestamp. This approach allows querying time-dependent tables very fast. In previous works [22,23], Timescale was compared with other time-series databases, demonstrating competitive performances on all the traditional queries. Working on millions of rows, Timescale offers up to 20× higher ingest

[3] https://postgis.net/.
[4] https://www.timescale.com/.

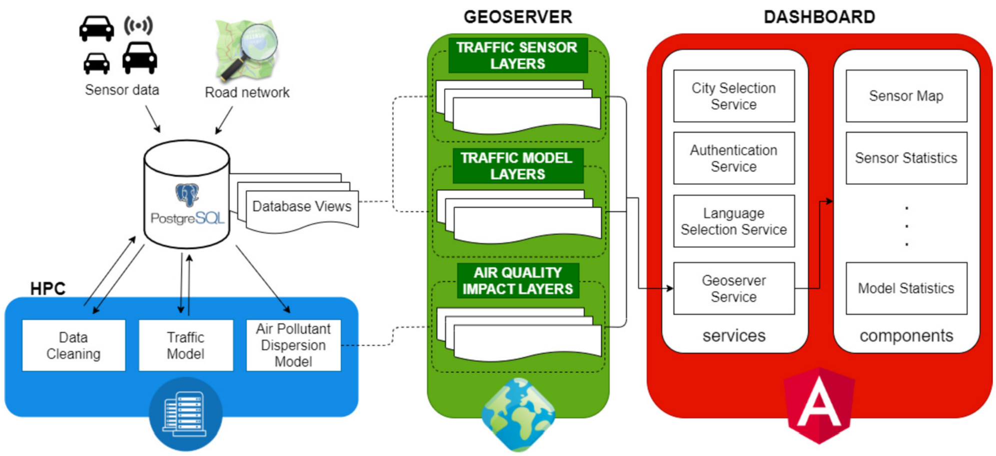

Fig. 1. Data flow and architecture of the Trafair framework.

rates, at the same time supporting time-based queries to be even 14,000× faster.[5]

The structure of our data model consists of 17 interconnected tables.[6] The data stored are related to the road network, the sensors (characteristics, positions, and measurements), the traffic model (configuration and output), the anomaly detection algorithm (configuration and results), and the statistics and trends of the traffic sensors measurements, the output of the two models: the air pollutant dispersion model and the traffic model.

The road network, the position of the sensors, and the neighborhoods of the city are represented with different geometries thanks to PostGIS extension. Several functionalities are exploited to perform spatial queries like obtaining the roads within a neighborhood or identifying in which road a sensor is located.

Timescale is exploited in the sensor data, the output of the traffic model, and the anomalies. These tables store a big amount of data, and their size grows quickly over time. Assuming that one sensor makes a measurement every minute, the total number of observations in one day will be 1,440, and approximately 43,200 in one month. If we suppose to monitor 400 sensors, in 2 years these sensors will produce more than 400 million records. Running a test on a table containing this amount of records, with and without the use of Timescale, we discovered that with Timescale we save from 50 to 99% of the time for each of the main queries performed.[7]

In the following, the storage of the road network and the acquisition of sensor data are described.

4.1.1. Road network

The road network is a key-element in traffic monitoring. In this context, Open Street Map[8] (OSM) is a valid resource that provides complete information about the road network for all the cities in the world and allows to filter and download the up-to-date geospatial data based on the use case. In addition, the spread of OSM in the world helps make our implementation reproducible in other cities. Other existing geospatial services such as Google Maps, Azure Maps, ArcGIS do not allow exporting raw map data, but only customize maps to include in other apps. This was not the scope of our use case; indeed, we want to export information, such as the geometry, related to each road in the area of interest in our data platform. Besides, the other services allow limited free use of data, requiring payment for a subscription later and imposing restrictions on the data usage. On the other hand, OSM is completely free, in accordance with the purpose of our project to use only open-source services. Another reason for choosing OSM is that the employed traffic model (Section 4.3) provides ready-to-use packages to import the road network from OSM.

The most important data for our scope are the road name and its geometry, the number of lanes, the maximum speed allowed, and the roadway directions. The number of lanes and the direction are helpful to geolocate the traffic sensors. The information in OSM is organized in ways, nodes, and relations. Among the ways, there are the roads which can be of different types: primary, secondary, pedestrian, cycle way, and others. While the nodes represent the points over the ways and the relations are used to model geographic relationships between objects like junctions and traffic restrictions.

In our data model, we ingest data related to roads and nodes. OSM data has been queried by using the Overpy Python library,[9] which allows filtering data based on the type of element (way, node, or relation) and the attributes associated with each element. The query can be limited to a specific area defined by the GPS coordinates. The obtained data are then stored in the above-mentioned data model. The traffic model (described in Section 4.3) exploits the OSM road network and converts it into a collection of road arcs. Consequently, the model output refers to the road arcs. Thanks to the PostGIS spatial functions, we manage to associate the model output to the OSM road network.

The urban area of a city is split into a collection of smaller sub-areas or neighborhoods according to the geospatial information coming from the geodata portal of the city of Modena.[10] These neighborhoods are represented as polygons in the Trafair database. The visualization obtained exploiting neighborhoods is described in Section 5.3.4.

4.1.2. Sensor data

There are several types of traffic sensors: induction loops detectors, video cameras, preformed loops detectors, magnetic detectors, pressure detectors, radars, and microwaves. The main data that these sensors are able to provide are the number of vehicles in the time interval (traffic flow) and their average speed. Consequently, the views implemented in TTD are based on these data and the information derived from them. In some cases, traffic sensors are also able to distinguish between vehicle types.

[5] https://blog.timescale.com/blog/timescaledb-vs-6a696248104e/.
[6] The database structure is depicted at https://trafair.eu/database_structure.html.
[7] A detailed description of this test is available at https://trafair.eu/timescaleperformance.html.
[8] https://www.openstreetmap.org/.
[9] https://pypi.org/project/overpy/.
[10] http://www.sistemonet.it/.

For each sensor, we stored a couple of GPS coordinates (i.e. a point in PostGIS) and the position within the OSM road network (calculated through the spatial functions of PostGIS).

The number of vehicles counted, and their average speed are stored as measurements provided by a sensor in a time interval, related to a vehicle type. The observation rates can have different values, according to the sensor and its configuration. Visualizations of sensor positions and measurements are reported in Section 5.2.

4.2. Data cleaning

Data cleaning is an effective way to improve sensor data quality. It aims to detect and eliminate data errors originated from the sensor measurements. We performed data cleaning in two steps: by filtering data that has an unreal correlation between flow and speed, and by applying an anomaly detection algorithm. The filter is applied as soon as the measurements are ingested into the database, flagging each measurement. While the anomaly detection algorithm is applied on filtered data. Anomalies are stored in the database, and excluded from the traffic model input. While the filter is applied to the measurements of the sensors as they are, the anomaly detection algorithm is applied to filtered sensor data aggregated by 15 minutes. This time interval can be customized, we chose 15 minutes since it is the same aggregation used by the traffic model (Section 4.3). Our implemented anomaly detection algorithm exploits the Seasonal-Trend Decomposition using Loess (STL) and the study of the Interquartile Range (IQR) on the remainder component of the time series. A detailed description of our approach for data cleaning is provided in [24]. An improved version of this approach has been explained in [25]. In this case, the data cleaning process, that previously contained only anomaly detection, is ameliorated with an anomaly classification phase where anomalies are classified into sensor faults and unusual traffic conditions; then, only observations classified as sensor faults are removed from the input of the traffic model.

4.3. Traffic model

Spatio-temporal data provided by sensors refer to fixed points in the city area. To gain a realistic insight into the traffic conditions in the city, widespread information is needed. For this reason, a traffic model is employed to reconstruct traffic flows and speeds in the whole urban network. The employed traffic model is Simulation of Urban MObility (SUMO),[11] a microscopic, space-continuous, and time-discrete simulator [26]. SUMO was configured to generate the routes of the vehicles starting from traffic sensors data. This model, described in [27,28], produces data about vehicle counts and their average speed in every road lane of the urban area. SUMO employs an own road network. We created this SUMO network by ingesting the road network described in Section 4.1.1. In the SUMO network, a calibrator is placed near each traffic sensor. Calibrators are SUMO objects that act like virtual traffic sensors, calibrated considering the real measurements of the on-road sensors. They aim to produce the expected traffic flow, i.e. the number of vehicles counted by the sensor associated with that calibrator in a defined time interval.

The model interacts with the Trafair database through Python scripts. We run daily simulations, and real-time simulations. The daily simulation simulates the 24 hours and starts at the end of each day. The real-time simulation is obtained as a collection of parallel simulations. Every 15 minutes a simulation based on the traffic sensor data of the last 3 hours is performed and combines traffic data with random and predefined vehicles' routes. Both simulation outputs are stored in the Trafair database. Information on the daily simulations is visualized in the "Traffic model history" view of the dashboard, as shown in Section 5.3.5. Information obtained simulating the last 15 minutes is visualized in the "Real-time traffic" view of the dashboard as described in Section 5.3.1.

4.3.1. Traffic data analysis

Urban traffic flows are spatio-temporal data series. Time series analysis concerns analyzing the evolution of values across the time dimension [29] to identify some trends that characterize data. In [30], an in-depth investigation of traffic flows obtained from our model was performed to discover trends. The Dynamic Time Warping (DTW) [31] is a well-known method to find the optimal alignment between two sequences, typically, time-dependent sequences. We employed DTW to evaluate the distance between two simulations. This distance is evaluated 'lanewise' for each road lane and then averaged on the whole map obtaining the mean DTW. While observing the mean DTW distance between several daily simulations, two simulations of the same day of the week generally have a reduced DTW distance. Moreover, each daily simulation was aggregated on the spatial dimension to obtain a unique time series: the trend of the average traffic flow evaluated on the whole urban area for each timestamp. A comparison between the obtained daily simulation time series underlines a similar trend for the same day of the week. Therefore, the daily simulation time series obtained for each day of the week were averaged to obtain the average day of the week trend.

In Fig. 2, the model output was elaborated to obtain the hourly sum of the vehicles circulating in the whole urban network for each day of 2019. Then, the distribution of the values of each day of the week has been represented as a boxplot. Comparing the boxplots, several observations emerge. The median of traffic flow during weekdays is similar, but the maximum value changes significantly. Friday and Monday have a higher median than the other days; the IQR (the difference between 75th and 25th percentiles) is similar on weekdays, and it is reduced on Saturday, this means less variability in the hourly flow; Sunday instead has an IQR similar in width but 75th and 25th percentiles have both lower values than on weekdays; during the weekend there is no evident peak, the maximum value is nearer to the median. On the right side of Fig. 2, the hourly trend of each day of the week is represented: the curve is obtained averaging all the available simulation outputs of 2019 for the given hour and weekday. It can be observed that Saturday has fewer vehicles in the morning hours but the number of vehicles during the afternoon is similar to weekdays. As expected, the number of vehicles counted during Saturday and Sunday nights is higher than on weekdays. Mondays are the most congested weekdays during the morning hours, but Wednesdays and Fridays are more congested in the afternoon and evening hours. All these observations highlight the necessity of studying each day of the week singularly. In Fig. 3, a similar analysis on the months of 2019 shows that also different months have different traffic flow value distributions and need to be studied individually.

Based on these analyses, a mean traffic flow value is calculated for every road arc and every 15 minutes interval, considering all the simulations performed in the given month on the specified day of the week. In this way, the spatial dimension is preserved, and the temporal dimension is aggregated on the weekday or the month. The result is a collection of geolocated time series, one for each road arc, for each day of the week, and each month of the year.

Another aggregation was used to further investigate the relation between traffic and seasonal weather. Traffic flow values were averaged for each season. Moreover, holidays can affect the traffic significantly. For this reason, days of the year 2019 were catego-

[11] https://sumo.dlr.de.

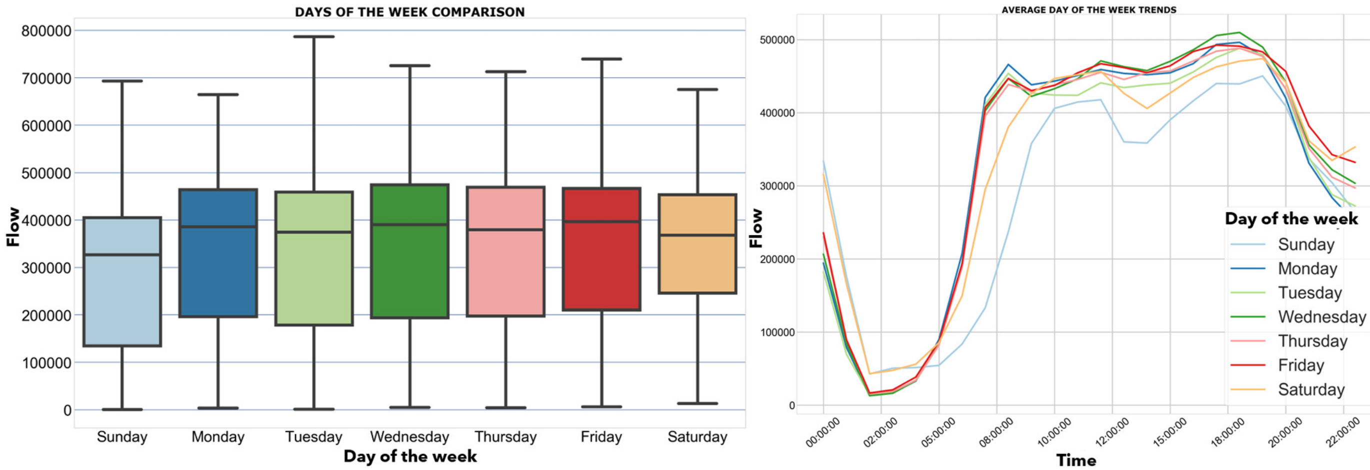

Fig. 2. Hourly traffic flow for each day of the week in the urban area of the city of Modena in 2019: distribution boxplots and trends.

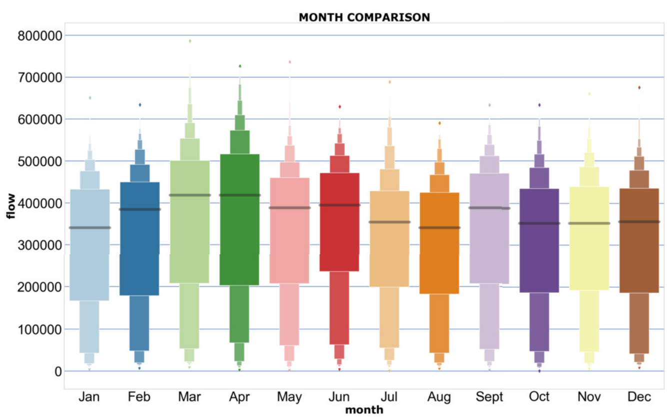

Fig. 3. Distribution of the hourly traffic flow for each month in the urban area of the city of Modena in 2019.

rized into two types of days: working day, or holiday. A mean value is calculated for every road arc and every 15 minutes, considering all the simulations performed in the given season for each day type. The resulting aggregation is collected in the Trafair database and the obtained visualizations are described in Section 5.3.2.

4.4. Air pollutant dispersion model

As a matter of fact, vehicle emissions strongly affect urban air quality. For this reason, an air pollution dispersion model was employed to predict the NO_x concentration in the urban area based on weather conditions, traffic flows, vehicle fleet composition, building shapes, and additional emitting sources (e.g. domestic heating, energy consumption, industrial combustion, and waste management).

Given the urban vehicle fleet, the NO_x emissions derived from traffic flows are evaluated by using a modified version of the R package VEIN (Vehicular Emissions INventories) v0.5.2 [32]. This package, described in [33], takes as input the simulated traffic flow and exploits a series of functions to automatically compute the total NO_x emissions for each road of the network. Emissions are evaluated considering the fleet composition of the city and the emission factors suggested by the European Environmental Agency [34]. Knowing the quantity of emitted particles at street level is not enough to provide pollutants concentration in the whole urban area: firstly because the emitted particles do not stay still but move in the air all around the city, and secondly, because the concentrations of pollutants do not depend only on traffic, but also on other sources. For this reason, the open-source simulation software Graz Lagrangian Model (GRAL) [35] is employed to simulate how emitted particles move in the air considering winds, weather conditions, and the shape of the building in the urban area. Moreover, GRAL takes into account also other emission sources (e.g. house heating) and produces a dispersion map that shows NO_x concentration values on an urban grid of 2-4 meters for every hour. The GRAL model generates 24 maps, one for each hour saved as GeoTIFFs. GeoTIFFs represent the concentration of NO_x in the urban area for the given hour. We investigate different scenarios: for each season of the year, we simulate a weekday and a holiday. Traffic flows are obtained averaging the weekdays and holidays traffic flow simulated in 2019 for each season. Then, weather conditions are defined as the typical weather conditions in that season of the year and are the same for the holiday and the weekday. Moreover, since the goal was to compare air quality conditions derived from different vehicle fleet compositions, we simulate each season weekday and holiday for different types of vehicle fleet compositions. The dashboard allows the user to visualize and compare the NO_x concentration derived from each one of the vehicle fleets in each season weekday and holiday (Section 5.4).

4.5. Data transformation

The database collects a huge amount of data that grows every day. Not all the available information needs to be visualized in the dashboard. To convey a good insight into the traffic conditions of the city, we have selected a pool of relevant visualizations that satisfies the requirements stated in Section 3. Data has been aggregated in space and/or time and several statistics have been extracted. To ensure easy and fast information retrieval the elaborated data are stored into materialized views that are updated regularly through automatic processes. The flexibility of the designed architecture is implemented thanks to an intermediate layer (GeoServer) that enables the abstraction between the data layer and the application layer. The flexibility and reproducibility of the framework are guaranteed. Even if different data models are implemented, it is enough that the required GeoServer layers are implemented to be able to display them in the TTD.

GeoServer[12] is an open-source server for sharing geospatial data. In GeoServer, data coming from different sources can be published over the internet privately or publicly through files (raster or vector data), tables in databases, raster files. GeoServer creates a

[12] http://geoserver.org/.

middle layer between the data structure and the external applications. The materialized views are converted into GeoServer layers that can be queried as API. Through these layers, GeoServer enables the handling of temporal and spatial data using the standards of the OGC protocols. The WFS (Web Feature Service) offers a direct fine-grained access to geographic information at the feature property level. The WMS (Web Map Service) is used for delivering map images, it generates server-side maps and sends them to the client. Moreover, WCS (Web Coverage Service) is a WFS for raster data: geospatial information is received as “coverages”. Coverages are objects that represent space-varying phenomena. Moreover, since we are managing spatio-temporal data, each image or object we are working with refers to a timestamp. For this reason, the Image Mosaic plugin was used to create a collection of rasters that refers to adjacent timestamps (i.e. the same day). WFS was used to expose data for drawing graphs and info-graphics. Instead, WMS was used to create maps and WCS with Image Mosaic was employed to save and retrieve the dispersion model outputs (Section 4.4). Section 5 describes the visualization obtained querying this information.

5. Trafair traffic dashboard

A geospatial dashboard is a tool for displaying and visualizing geospatial data and a resource to support data mining and decision-making. Trafair Traffic Dashboard[13] was created with the aim of providing real-time traffic information, sensor measurements and derived statistics, traffic model historical data, and emerged trends and responding to the needs and requirements of the Trafair partners as stated in Section 3.

The dashboard can provide data coming from three different sources: the traffic sensor located in the city, the output of the traffic model, and the output of the air pollutant dispersion model in different conditions. The data coming from sensors are geo-located time series that refers to a point in space (the location of the sensor). The traffic model output instead is a collection of time series that refers to a road section lane, whose geometry is simplified as a segment.

The visualization of sensor data can help public authorities to understand the measured traffic flow and speed in a certain location and to perform sensor maintenance, discovering anomalous behavior. The traffic model output is generated from simulations based on the measurements of the traffic sensors. It provides traffic information in the whole urban area. Statistics and historical data can help to understand critical traffic conditions, the most congested areas of the city, and the traffic trends at different times of the day or periods of the year. The air pollutant dispersion model output is generated from simulations mainly based on the traffic emissions. The visualization of NO_x concentration derived from the air pollutant dispersion model can help the city council to compare different scenarios and better understand how vehicles fleet composition, seasons and holidays can influence the quality of the air.

As displayed in Fig. 1, the dashboard queries GeoServer layers to extract information displayed in views. In the following subsections, the implementation choices for the creation of the web application and the available views will be described.

5.1. Dashboard implementation

The dashboard is a web application implemented using Angular 7 application design framework and Typescript language. Two JavaScript libraries have been used to display graphs: D3,[14] and Chart.js.[15] D3 is a framework that enables proprietary representation and extraordinary flexibility in the creation of charts and graphs that support large datasets and dynamic behaviors for interaction and animation. Chart.js is a library that provides a selection of predefined charts that can be customized. Open Layer[16] instead is the library employed to display dynamic maps. Angular is a platform for building single-page client applications with a modular structure. Angular code is based on building blocks called NgModules: a collection of related codes into functional sets. These modules generate “components”. Each component is composed of several screen elements. The functionalities that are shared between different dashboard views are provided by “services”. Services are employed to share data, information, and functionalities between components. In our implementation, several services were defined to share functionalities and information. The created services can retrieve and modify the information shared between components.

The main services (as displayed in Fig. 1) are: (i) Authentication service: to share and modify data (permissions, status) concerning the user that are logged in; (ii) City selection service: to share and modify data about the selected city (our web application supports two cities); (iii) Language selection service: to share and modify the selected language (in our application the user can choose between Italian, Spanish, or English); (iv) GeoServer service: to share functionalities related to GeoServer interrogations.

5.2. Traffic sensors data visualizations

Traffic sensor data are collected inside the Trafair database and analyzed to produce statistics, trends, and sensors fault recognition. The available views show the sensors' positions, the latest measurements with anomalies, a comparison of sensors located in the same crossroad, and statistics. The majority of views regarding traffic sensor data in the TTD show traffic flow instead of speed. The reason is that the traffic sensors installed in the city urban areas are usually located near crossroads and traffic lights and the measured speed is less significant to evaluate traffic conditions.

5.2.1. Sensor map and measurements

The sensor map view allows the user to understand traffic sensor distribution in the urban area and easily recognize faulty sensors. As shown in Fig. 4, a marker in each sensor position is displayed. The markers are blue if the behavior of the sensor is normal, gray if it is not providing any observation in the last 24 hours, and red if the sensor is faulty. A sensor is considered faulty when one of the following conditions is verified: (i) the percentage of filtered observations in the last 24 hours is higher than 50% of the total number of provided observations, (ii) the percentage of anomalies detected in the last 24 hours is higher than 50% of the total number of provided observations. Clicking on a sensor marker, the user can observe the last 24 hours observations aggregated every 15 minutes, as displayed in Fig. 5. The view shows two time series: the number of counted vehicles, and their average speed. In both time series, the anomalous observations are highlighted with a red dot. In the same view, additional information about sensor conditions is displayed: the percentage of filtered observations (note that the filter is applied on not aggregated observations), and the percentage of anomalies (anomalies are detected on observations aggregated every 15 minutes). This page can help public administration members to check the sensor conditions. To

[13] The dashboard is available online at https://trafair.eu/trafficflow. Test credentials are: user: test, password: test.

[14] https://d3js.org/.

[15] https://www.chartjs.org/.

[16] https://openlayers.org/.

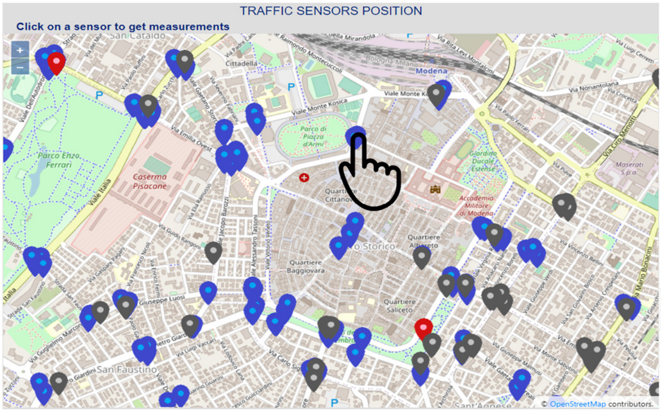

Fig. 4. Sensor map view. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

help them investigates the nature of anomalies, through the "show nearest sensor measurements" button, the users can visualize the current sensor time series of observations (in red) compared with the time series of the sensors placed in the same crossroad (in black), as displayed in Fig. 6. The described views satisfy the **R1** requirement described in Section 3.

5.2.2. Sensor data statistics view

For each sensor, a large amount of data is collected in the Trafair database, therefore statistics can help the user to understand the collected information. The values of traffic flow are aggregated to obtain daily views. Three different graphs are available:(i) total number of vehicle counted, (ii) percentage distribution of vehicles, and (iii) top 30 sensors. The first one is a stacked bar graph that shows the total number of vehicles counted for each day of the month and each period of the day. Six periods are defined in a day. This graph is useful to understand in which period the number of vehicles is higher and how this can change on different days of the week or months. The "percentage distribution of vehicles" graph shows the percentage of vehicles counted each day on the total number of vehicles counted on the whole month. This gives an idea about how much each day contributes to the total vehicle flow of the month. Finally, the "top 30 sensors" (Fig. 7) displays a histogram that shows the number of vehicles counted in a selected day for the 30 sensors that had the highest counts. The user can interact with the graph: clicking on one bar, representing the number of vehicles counted by a sensor, and the map on the left will show the position of that sensor. This view is useful to understand where the areas of the city with the highest concentration of vehicles are, and, eventually, to recognize sensors with a very high vehicle count that could be due to sensor faults.

The views allow the user to compare different days of the week, months of the year, and the same month in different years. The displayed graphs enable users to easily recognize a day that has a traffic flow different from the others and investigate the points of the city where vehicle counts were higher. The described visualizations satisfy the requirement **R2** established in Section 3.

5.3. Traffic model output visualizations

The traffic model simulation output contains information about traffic flow and speed for each road segment in the simulated area. The data are collected inside Trafair database and analyzed to obtain average weekdays and seasonal trends. The dashboard enables easy navigation through space and time to compare the different periods of the day, of the week, of the year, and different parts of the city.

5.3.1. Real-time traffic view

In this view, the city streets are depicted with different colors according to the traffic conditions evaluated by the real-time traffic model (described in Section 4.3). The map is automatically updated every 15 minutes. Each lane is depicted as a line colored according to the traffic density (vehicles/km). An example of this view, for the city of Modena, is displayed in Fig. 8.

The map is for information purposes. It allows users to immediately grasp if there is congestion in place and in what area of the city and, therefore, satisfies the **R0** requirement stated in Section 3. By clicking on the button located in the top-right of the page the colors of the road lanes will change according to the traffic flow (vehicles/hour) color scale.

5.3.2. Traffic model global statistics view

This view allows the user to visualize seasonal and monthly trends. After selecting the month or the season, the user can interact with the graph by removing some curves by clicking on their label in the legend, thus focusing his attention on a few. In Fig. 9, the trend of the weekdays of March 2019 in the city of Santiago de Compostela is displayed. To obtain this graph, the hourly sum of vehicles in the whole urban area was calculated for each day of the month and then, all simulations concerning the same weekday have been averaged. Moreover, in Fig. 10, the trend of weekdays and holidays in autumn of 2019 is shown. To obtain this graph, days have been classified as weekdays and holidays, then the hourly sum of vehicles in the whole urban area was calculated for each day of the selected season. Finally, the average curve was evaluated for weekdays and holidays. This view satisfies the **R3** requirement (as stated in Section 3).

5.3.3. Traffic model speed index view

This view allows the user to visualize the Speed Index (SI) profile in the whole urban area. The SI is obtained as the ratio between the average hourly speed and the speed limit in the road section. If the value is lower than 0.6, vehicles are moving slower than expected, this can happen if there is traffic congestion. If the SI value is higher than 0.6, vehicles are moving accordingly to the speed limit, and there is no congestion in that road lane. The speed index profile is generated for each day of the week of each month. The user can select the month and the day of the week. Then, he/she can use the scroll-bar on the map to select the hour of the day he/she wants to investigate (Fig. 11). Once selected the month, the day of the week, and the hour of the day, each road lane in the map is colored according to the value of its SI. The color scale was defined considering that if vehicles have an average speed that is around the 20% of the speed limit in that road section, there are traffic jams. If the percentage is between 20% and 40%, there are slowdowns; between 40% and 60%, the traffic is lowly congested; between 60% and 80%, the traffic conditions are normal; above 80%, the free flow is reached. The view was created to satisfy the **R3** requirement stated in Section 3.

5.3.4. Neighborhood trend view

The traffic flow trend of an average day of the week, as displayed in the Traffic model statistics view described in Section 5.3.2, summarizes the traffic flow in the whole urban area aggregating on the spatial dimension. However, citizens are also interested in knowing the condition of traffic in the place where they live or work. Moreover, decision-makers at City Council need to explore neighborhood-level statistics and see how traffic in their communities has changed over time. To further investigate the spatial dimension, the neighborhood trend view shows a trend for each neighborhood of the city (Fig. 12). The exploration of neighborhood-level statistics is available thanks to the geospatial information coming from the geodata portal of the city of Modena,

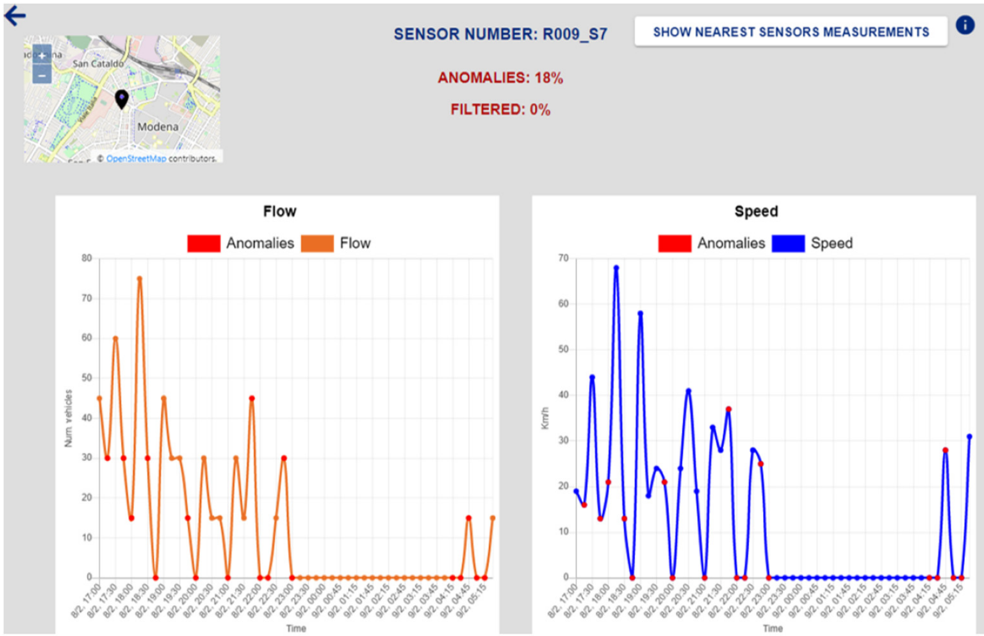

Fig. 5. Measurements and anomalies for the selected sensor.

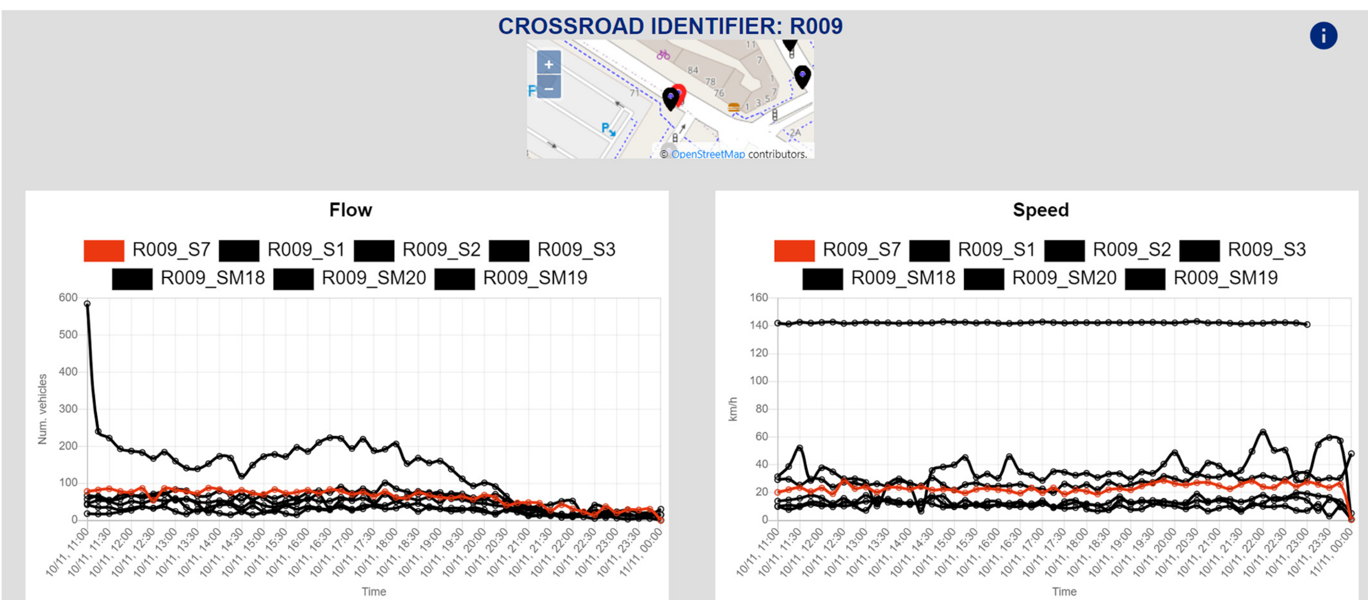

Fig. 6. Comparison between the time series of the selected sensor and the time series of sensors located in the same junction.

as described in Section 4.1.1. The user can select the day of the week, then a graph with a curve for each city's neighborhood is displayed. The trend is evaluated as the average traffic flow of the road sections located inside the neighborhood area. The decision to average the traffic flow instead of summing up the number of vehicles enables to compare neighborhoods with a different extension and, as a consequence, a different number of road sections. This view integrated with data regarding the number of schools, churches, residential areas, industrial areas can help to better understand the activity-based routes and plan traffic management. This view satisfies the **R3** requirement stated in Section 3.

5.3.5. *Traffic model history view*

In this view, the overall traffic flow trend simulated by the model for the selected year is displayed. The historical data are obtained as a collection of daily simulations performed during the year. The user can interact with the graph by selecting a rectangular area to focus on (Fig. 13), then the graph is zoomed according to the selection. The curve is obtained summing up all the vehicles in the whole urban area aggregated per hour. The view allows the user to see the weekly trend and the monthly and seasonal trends of the whole selected year satisfying the **R4** requirement presented in Section 3.

5.3.6. *Annual average daily traffic volume view*

The most commonly used measure of traffic is the annual average daily traffic (AADT) volume. Theoretically, AADT is defined as the average 24 hours traffic volume at a given road lane over a full 365 days/year [36]. For each road section x, AADT was evaluated as:

$$AADT_x = \frac{\sum_{i=0}^{D}\sum_{j=0}^{24} flow_{x,j,i}}{D}$$

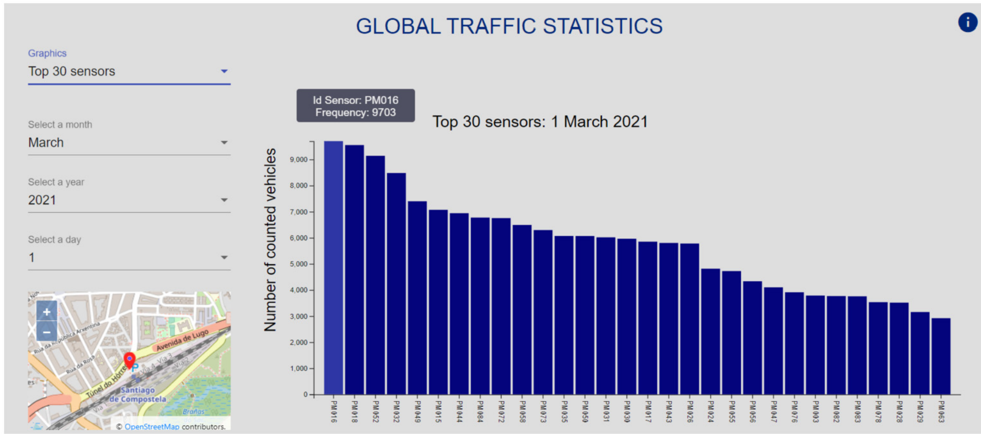

Fig. 7. The visualization of the 30 sensors with the highest vehicle counts on the 1st of March 2021 in the city of Santiago de Compostela. The map displays the position of the selected sensor.

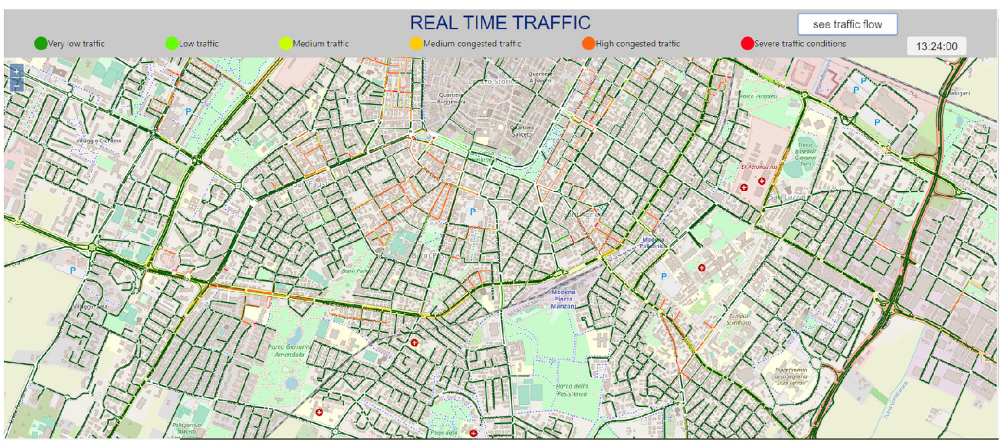

Fig. 8. Real-time traffic view of the city of Modena based on traffic density at 13:24 UTC on the 30th of July 2021.

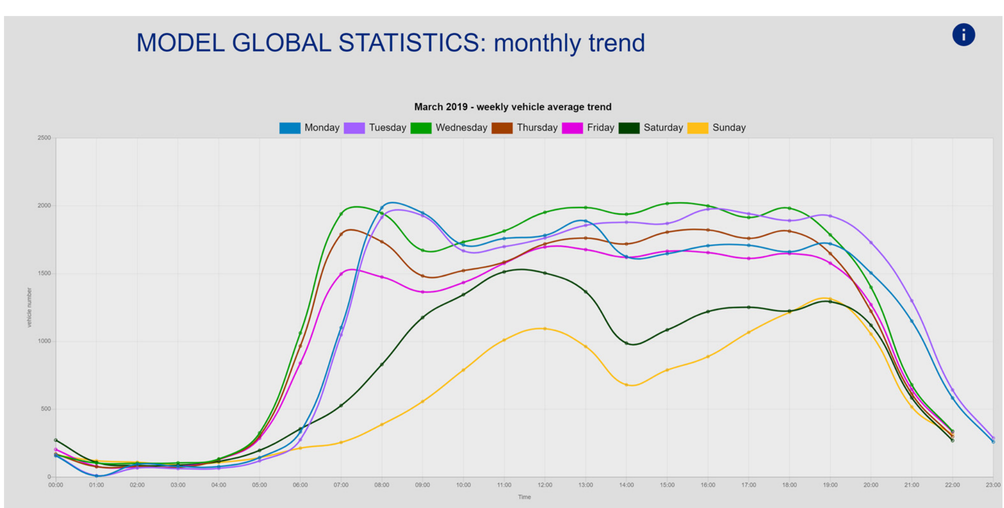

Fig. 9. Traffic Model Statistics in the city of Santiago de Compostela: average traffic flow trend for each day of the week of March 2019.

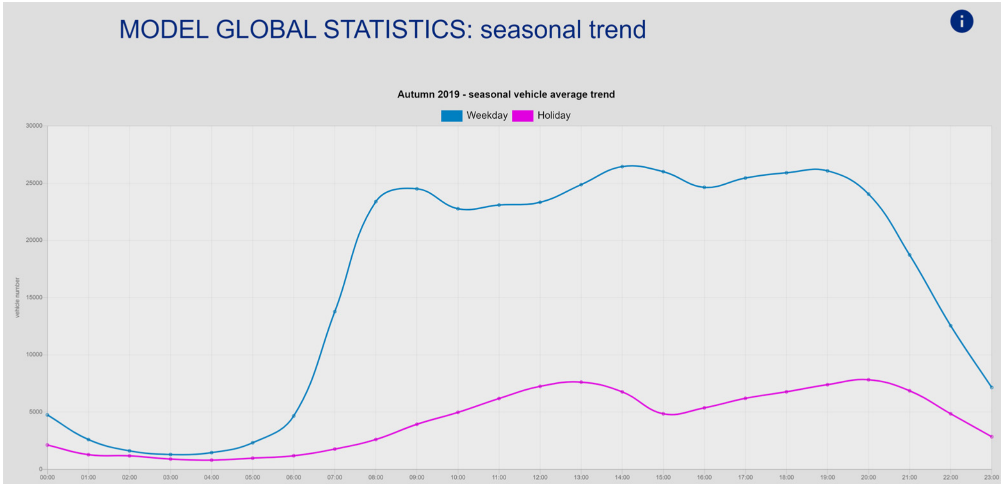

Fig. 10. Traffic Model Statistics in the city of Santiago de Compostela: average traffic flow trend for working days and holidays in the autumn season of 2019.

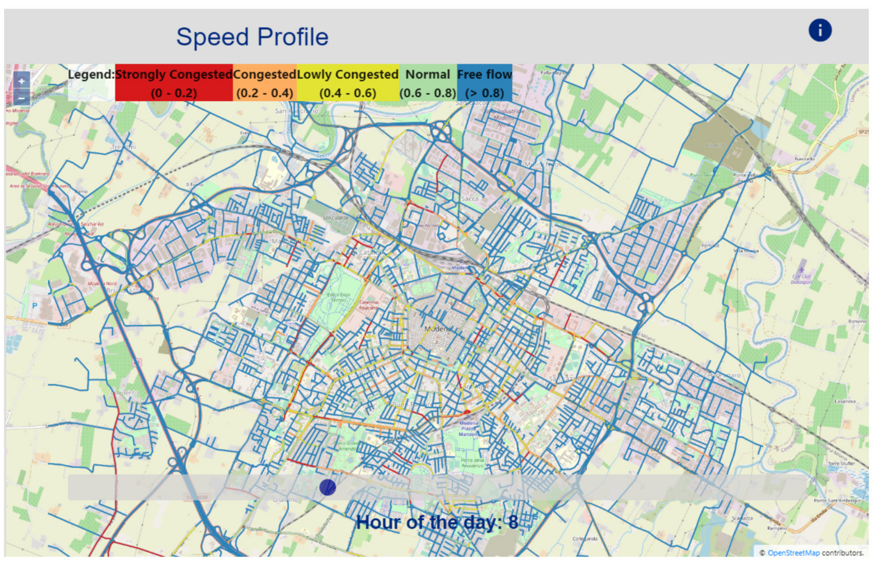

Fig. 11. The Speed Index profile in the city of Modena on an April Wednesday at 8 AM. The colors of the road lanes refer to the measured SI. The lower is the value of SI, the slower is the traffic in that lane.

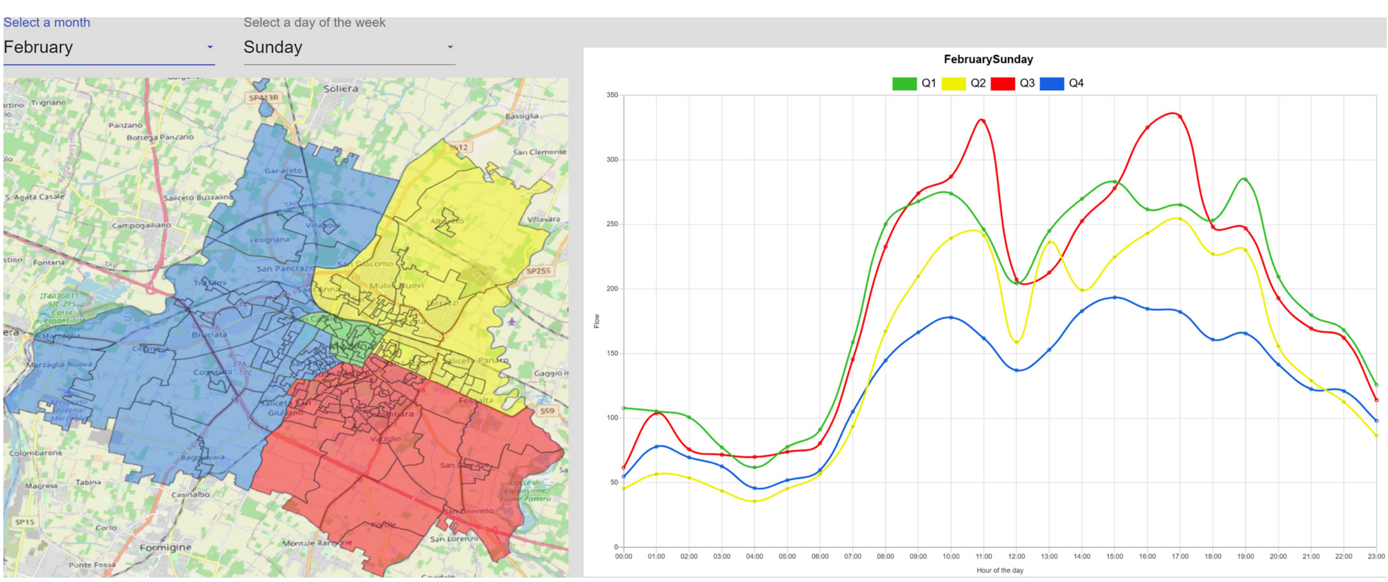

Fig. 12. Average traffic flow trend of road sections located in different neighborhoods of the city of Modena on a February Sunday.

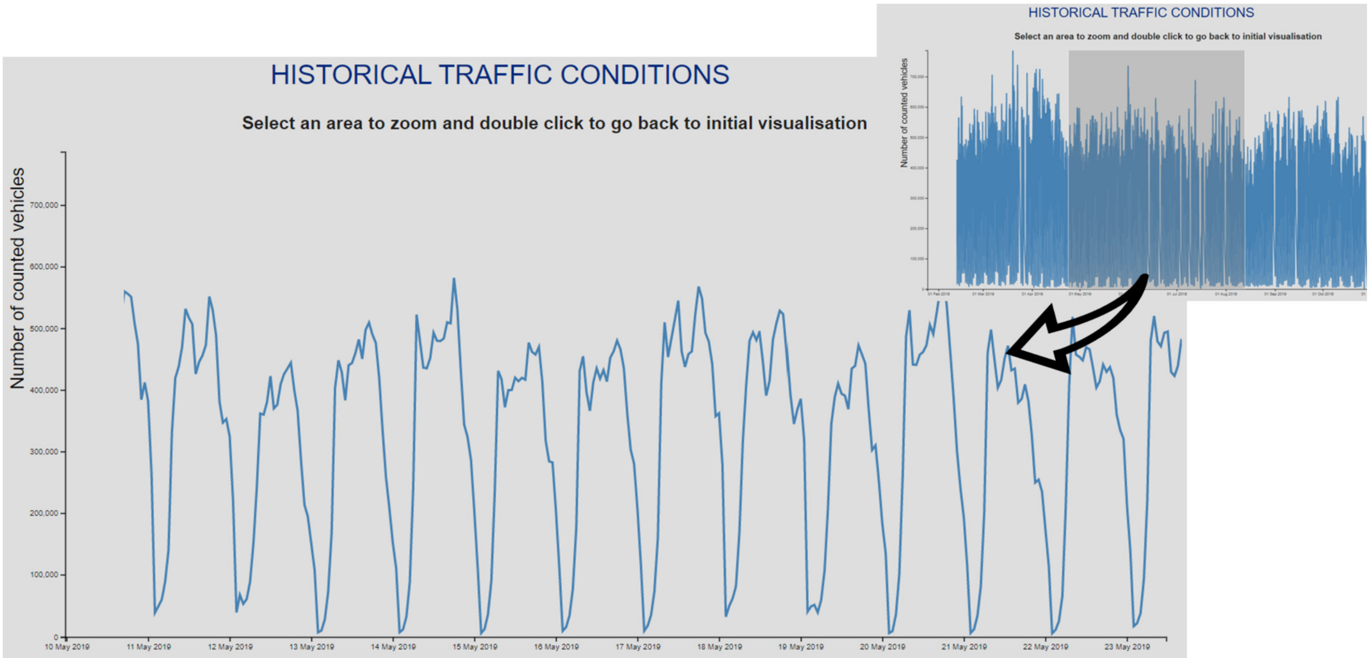

Fig. 13. Historical traffic flow trend of 2019. In the smaller view, the user is selecting a rectangular area. In the bigger view, the graph obtained zooming the selected area is displayed.

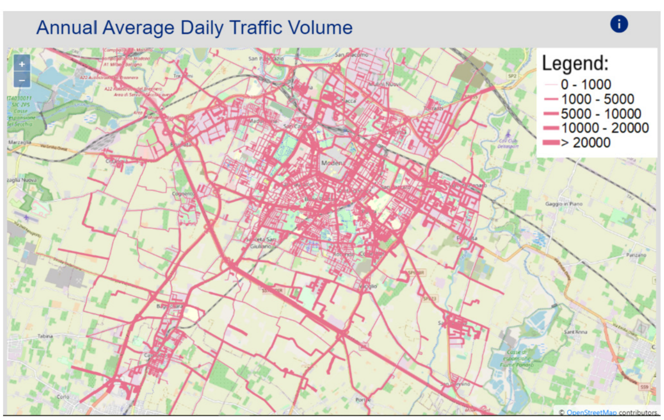

Fig. 14. AADT map of 2019 for the city of Modena.

where D is the total number of simulated days in the considered year, and $flow_{x,j,i}$ is the observed traffic flow (veh/hour) in the road lane x for the j^{th} hour of the i^{th} day of the year. AADT enables to easily compare traffic volumes of different years on different roads and understand how they have changed over time. The idea is to aggregate traffic volume on the time dimension and then explore the spatial dimension of the data. The dashboard shows a map where the road lanes with the highest AADT are represented with a thicker line (Fig. 14). The employed scale was defined considering the urban context with the absence of big highways. The user can interact with the map zooming and moving in the area. The roads with the highest traffic volume can be easily detected, helping the municipality in urban planning. Once selected the year, the AADT map of the selected year will be displayed. In the left part of the view, a map shows the road lanes that had a higher traffic volume in 2019 colored in green and the ones with higher traffic in 2020 colored in red (Fig. 18). The size of the line depends on the absolute value of that difference. This view allows easy comparison of the two years. The requirement **R3** described in Section 3 is satisfied by this visualization.

5.4. *Air quality impact visualization*

Sustainable smart cities are concentrated upon mitigating air pollution that can be achieved also by limiting vehicular traffic. Most unhealthy air quality conditions occur in city centers, mainly caused by traffic-related NO_x emissions originating from cars. In this scenario, understanding the influence of the vehicle fleet composition on NO_x concentration in the urban area is crucial. Each city defines a set of different scenarios in line with its sustainable mobility plans. The vehicle fleets are described with a pie chart that allows the user to visualize the percentage of each fuel type. Then, using a drop-down menu the user can select the season and type of day he/she wants to examine and the NO_x concentration will be displayed in the three maps at the bottom of the page. The map at the top instead shows the traffic flows used as input to the dispersion model. The user can interact with the visualization using the scroll bar and selecting the hour of the day he/she wants to inspect. All the maps will be updated according to the select season, day type, and hour of the day. This view ensures an easy comparison between the different vehicle fleets and can help to understand also the impact of changes in the vehicle fleet in different seasons. Fig. 15 shows a comparison between air quality on a winter weekday at 6 PM with the actual vehicle fleet and the other two possible vehicle fleets. The color scale helps the user to understand the entity of the actual problem and to investigate the efficacy of the increase in electric and hybrid vehicles. These visualizations satisfy the **R5** requirement stated in Section 3.

6. Case studies

TTD has been successfully implemented in two cities: Modena (Italy) and Santiago de Compostela (Spain). These cities have different geographical characteristics and adopted a bit different technologies and strategies, as will be discussed in the following sub-sections. The first release of the dashboard took place on April 30^{th}, 2020 and the data displayed refer to the period from 2019 till today. Thanks to the dashboard, it was possible to analyze and compare more than two years of traffic data for both cities. Below,

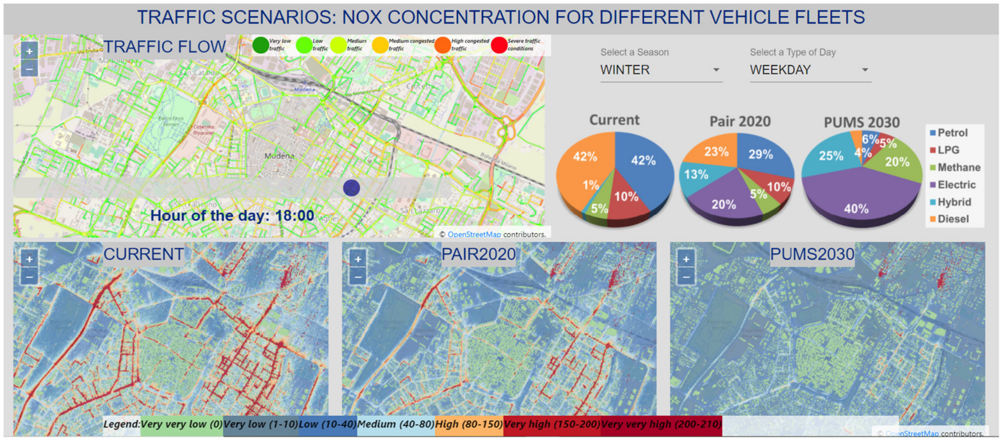

Fig. 15. Comparison between hourly traffic flow and NOx concentrations derived from different vehicle fleets on a winter weekday at 6 PM in the city of Modena.

some interesting views are reported and discussed for each of the two cities.

6.1. Modena

The city of Modena has an area of 183 km^2 and 184,727 inhabitants. In Modena 400 sensors are available, most of which are located near traffic light intersections, while others are placed on provincial or state roads in the main access points to the city. The Municipality of Modena and Lepida S.c.p.A. is in charge of the management of these sensors. The traffic sensors are induction loop detectors, placed under the surface of the street. Some sensors provide a measurement every minute whereas the others every 15 minutes. The sensor data collection started in November 2018 and is currently running day by day. From November 2018 till now (11^{th} March 2021), we have stored in our database 213 GB of sensor data for the city of Modena. Sensor data are successfully used by the traffic model implemented in Modena [28] which is executed on an HPC platform.[17]

TTD is used very frequently by both citizens and public administrators. Public administrators have no limitation in accessing the views. For example, they can consult data regarding sensors positions and last 24 hours measurements. These features help them to recognize faulty sensors ensuring sensor maintenance. Citizens, instead, can only access sensor data statistics and traffic model views.

The “sensor map and measurements” (Section 5.2.1) view was very useful to identify faulty sensors and thus to exclude them from the input of the traffic model. In the “sensor data statistic” view (Section 5.2.2) the “top 30 sensors” graph was strategically important to recognize sensors with suspicious behavior. For example, in Fig. 16, the number of total vehicles counted by sensors R118_SM117 and R118_SM118 compared with all the other sensors appears to be 7 times higher. This high count led us to take a deeper look at the sensor measurements and we realized that the number of vehicles counted could not be reasonable compared to the speed observed. This and other similar considerations that we were able to draw from the visualizations already available in the first version of the dashboard led us to define the filter and apply it to all data as described in Section 4.2.

The “total number of vehicles counted” and “percentage distribution of vehicles” graphs help public authorities to better understand traffic in the city. In Fig. 17, it can be observed, as expected, that the total number of vehicles on weekends and holidays (highlighted with black rectangles) is lower than the total number of vehicles on weekdays. However, the number of vehicles circulating during night hours of that weekends and holidays is higher. The number of vehicles in the morning peak hours on Sundays is the lowest among all days. This visualization can suggest additional insights: Friday appears to be the weekday with the highest traffic flow due to a high number of vehicles circulating during the afternoon.

The city of Modena is divided into four neighborhoods. The dashboard allows citizens and public authorities to visualize the traffic evolution during the day of the week in each of them. In Fig. 12, the average trend of the four neighborhoods of the city of Modena on a November Friday is displayed. The graph shows that the central area of the city (Q1) has a more evident peak in the morning hours. Then, the number of vehicles decreases during the afternoon. The north-east area of the city (Q2) instead has its peak in the afternoon, around 5 PM. The west part of the city (Q4) is the one with the lowest average traffic flow. The traffic conditions in these areas considerably change in different months and days of the week.

The “Air quality impact” visualization described in Section 5.4 helped public authorities to understand the effect of differently composed vehicle fleets in the city. For the city of Modena three fleets are investigated: (i) CURRENT: the actual vehicle fleet composed mainly of petrol and diesel alimented vehicles, (ii) PAIR 2020: the vehicle fleet suggested by the Integrated Air Plan of Emilia-Romagna region with a higher percentage of electric and hybrid vehicles, and (iii) PUMS 2030: the desirable future vehicle fleet, announced in the Urban Sustainable Mobility Plan, to be reached in 2030 with a majority of electric and hybrid vehicles. Fig. 15 shows an evident reduction of the NO_x concentration generated by the current vehicle fleet with the PAIR 2020 vehicle fleet and an even further decrease with PUMS 2030 suggested fleet composition.

The “Annual Average Daily Traffic volume” (AADT) view enables users to grasp general insight of the traffic condition in the city. Fig. 18 shows a comparison AADT map between 2019 and 2020 for the city of Modena: the bigger is the absolute value of the difference, the thicker is the line. It can be observed that in the city of Modena in 2020 there was a general reduction in average daily traffic. Another annual comparison is reported in Fig. 19, where the annual trend of 2020 and 2019 is discussed through the “Traffic model history” view. This visualization highlights the reduction in traffic due to the outbreak of the SARS-CoV-2 pandemic and the consequent restrictions to mobility.

[17] The HPC platform is a Linux based heterogeneous cluster, with an Infiniband FDR low latency network (56 Gb/s). The employed node consists of 8 Xeon 20-Core 6230, 2.1 Ghz, and 1.5 TB RAM.

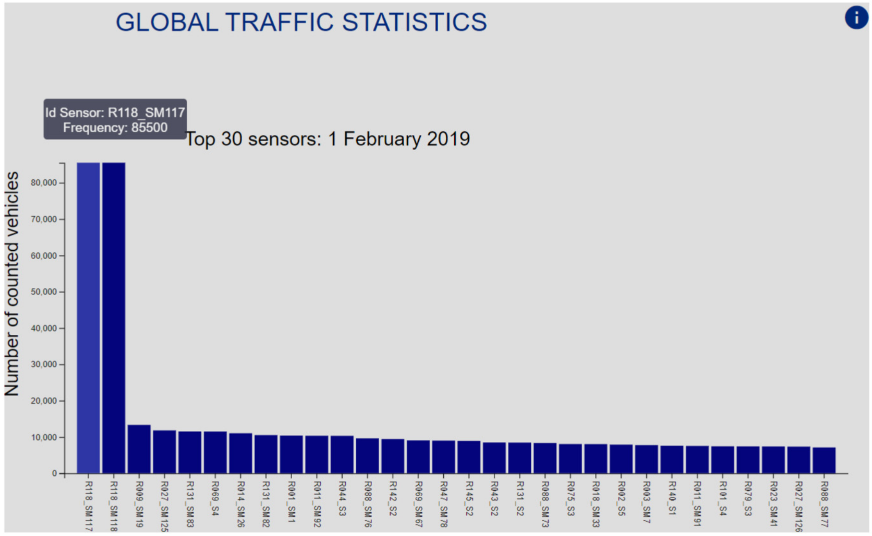

Fig. 16. "Sensor data statistic" view that shows the 30 sensors with the highest vehicles count on the 1st of February 2019.

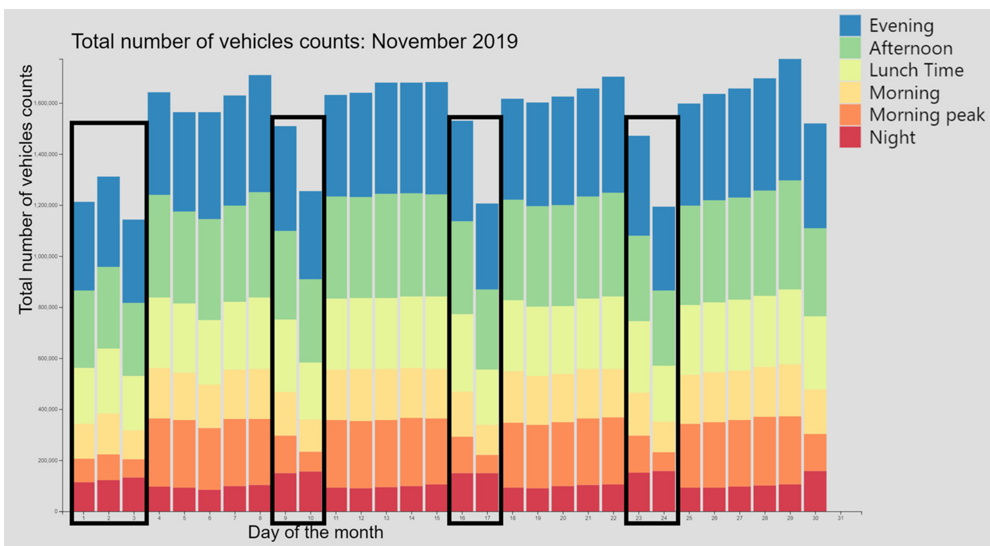

Fig. 17. Total number of vehicles for each day of November 2019.

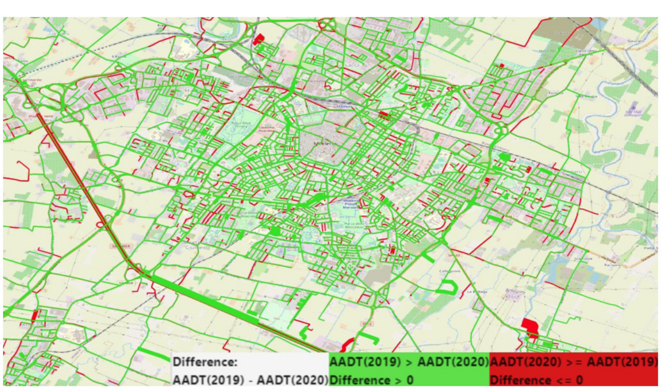

Fig. 18. A city map showing the difference in AADT between the years 2019 and 2020.

6.2. Santiago de Compostela

TTD is efficiently employed also in the city of Santiago de Compostela. Currently, the dashboard is not open to citizens, but it is reserved to the members of the City Council. Santiago de Compostela has 93,584 inhabitants and covers an area of 220 km^2. The models are executed on an HPC platform.[18] 74 traffic sensors (induction loop detectors) are located in the city. Each sensor supplies an observation every 5 minutes concerning the number of vehicles counted, while the average speed is not provided. The "sensor map and measurements" (Section 5.2.1) view shows the position of sensors. They are all displayed with the same color since no data cleaning is executed on these data. "Traffic model global statistics" view can help to understand the traffic flow trend in the city. In Fig. 9, a comparison between the different days of the week in March 2019 shows that all weekdays have a similar trend, however, Monday has a lower number of vehicles during the afternoon and the evening. On the weekend, the number of vehicles is significantly reduced. Similarly, in Autumn, the average weekday flow is more than three times bigger than the average holiday flow. The "total number of vehicle counted" graph, described in Section 5.2.2,

[18] The HPC platform is a Linux based heterogeneous cluster, with an Infiniband FDR low latency network interconnecting 317 computing nodes based on Intel Xeon Haswell processors. These nodes are able to provide a computing power of 328 TFLOPS, 44.8 TB RAM, and 1.5 PB disk capacity.

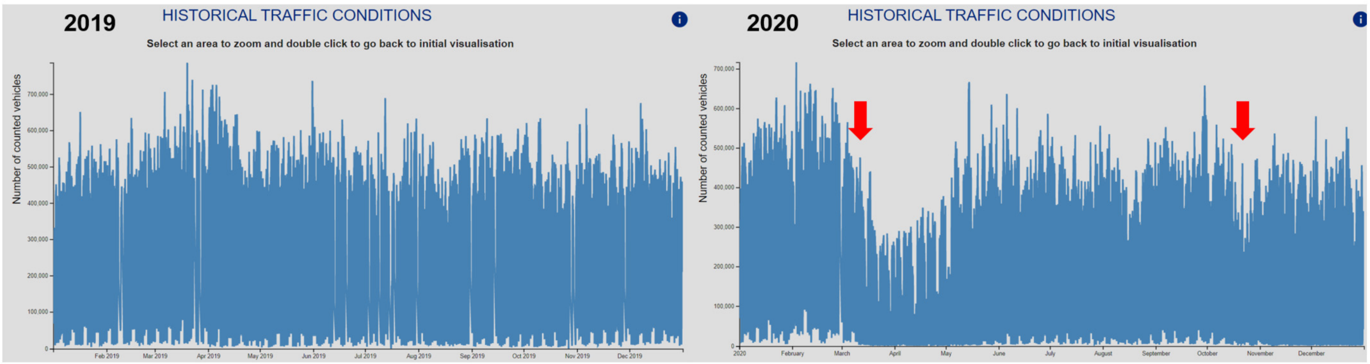

Fig. 19. Comparison between the annual trend of traffic flow generated by the traffic model in 2019 and 2020. The red arrows highlight the two pandemic waves in 2020.

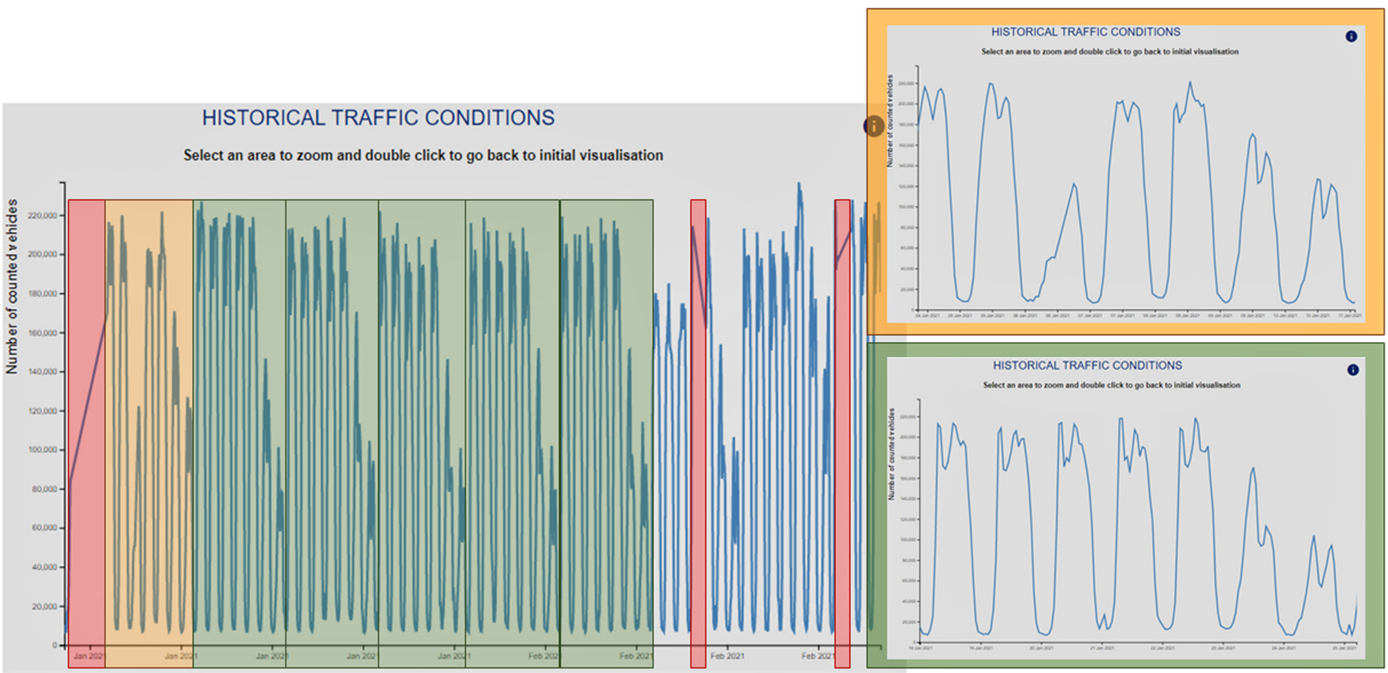

Fig. 20. Total number of simulated vehicles per day in 2021 in Santiago de Compostela.

is employed to investigate the effects on traffic of the second wave of the Sars-CoV-2 pandemic in Santiago de Compostela.

In Fig. 20, the “Traffic model history” view of the first months of 2021 is shown. In the graph that appears on the left of the Figure, the periods in which no simulations have been performed are highlighted in red, while in yellow and green 6 weeks of traffic simulations are spotlighted. The first week (in yellow), as shown in the graph on the top right, shows an irregular trend due to the holiday on January 6^{th}. On these days, the traffic is much lower than the average on a weekday and there is no peak in the morning hours. The following 5 weeks, highlighted in green, instead, have a similar weekly trend. As shown in the graph on the bottom right, weekdays have a similar trend among each other while on Saturday and Sunday the traffic flow is considerably lower.

In Fig. 21, November 2019 (on the left) and November 2020 (on the right) are compared. It can be observed that the total number of vehicles is significantly reduced in November 2020. Moreover, the restriction imposed by the Spanish government to observe a curfew during the night hours resulted in a sharp decreases in the number of vehicles circulating during the night period.

Besides, in Fig. 22, the traffic flow of March 2020 is shown. During this month the Sars-CoV-2 out broken in all of Europe. This graph displays the decrease in the number of vehicles due to the adoption of a strict lockdown in Modena on the 10^{th} March, and in Santiago de Compostela on the 14^{th}. The impact of the lockdown in the city of Santiago de Compostela was more rapid and the reduction of circulating vehicles more evident than in the city of Modena.

7. Discussion and limitations

The research conducted has highlighted that visualizing significant insights requires an analysis of the traffic-related data from different points of view. Traffic flow visualizations in semi real-time might reveal traffic congestion immediately, while real-time sensor data can detect abnormal behavior or even sensor faults. Traffic data statistics over the months or the years provide a clear understanding of similarities and dissimilarities among days, and the global number of traffic counts can be taken as an indicator of the mobility in a city. Simulated data can provide a view over the entire urban area and help in understanding different behaviors in different sub-areas of neighborhoods. Moreover, traffic flow simulations are the only effective means to understand the impact of traffic on air quality. It is only by simulating new hypothetical scenarios of the traffic that we can see what will be the impact of new traffic restrictions or limitations. On the other hand, these

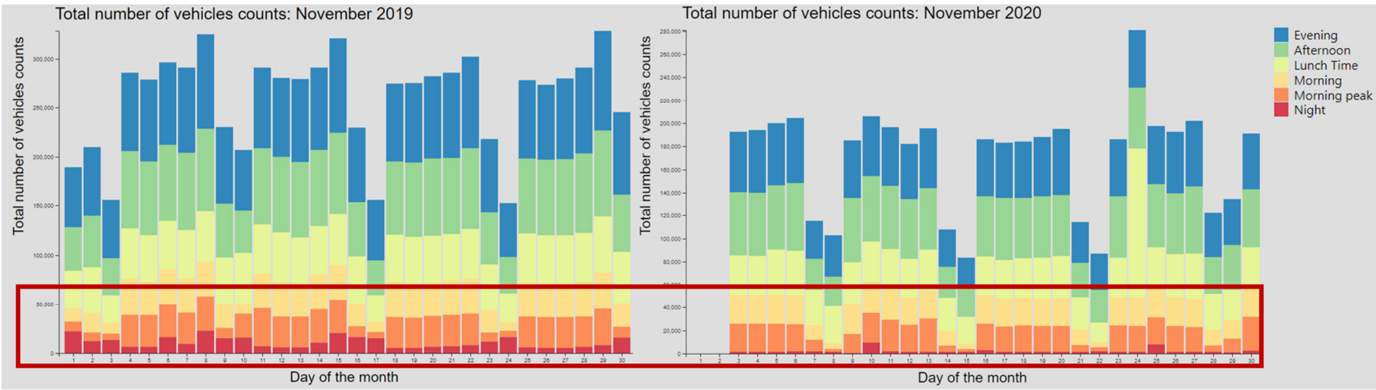

Fig. 21. Comparison between observed traffic flow in Santiago de Compostela on November 2019 and November 2020.

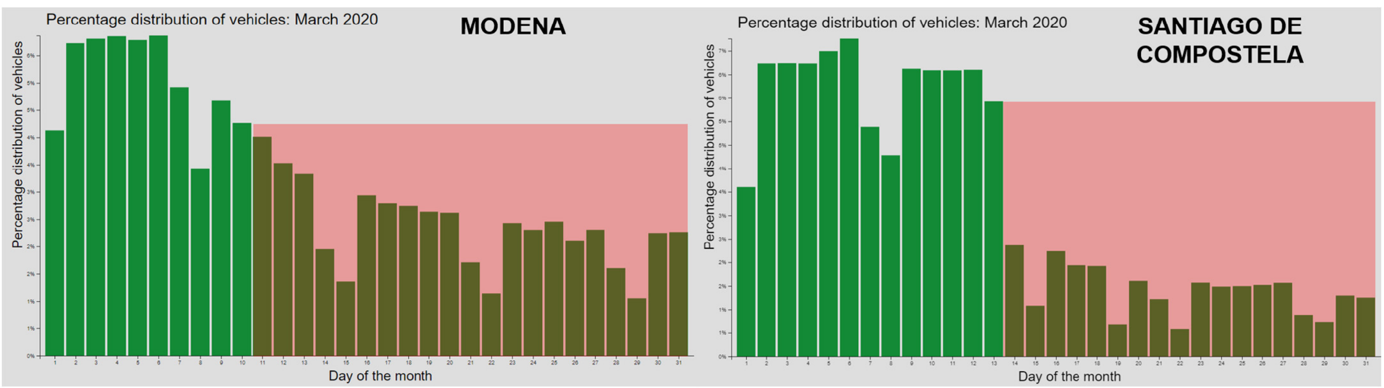

Fig. 22. Daily percentage over the total number of counted vehicles in March 2020 in the city of Modena and Santiago de Compostela. The lockdown periods are highlighted by red rectangles.

maps that describe the impact of traffic on air quality are raising the awareness of the citizens that work, live, or transit in that specific area.

Our approach provides insight into the spatial distributions of the data as well as the evolution over time. The available visualizations make it difficult to compare the same views referring to different periods. In the future, some additional visualizations will be integrated to enable an efficient comparison of the spatial distribution of traffic over time.

The use of GeoServer enhances the security of the system and ensures the flexibility of the designed architecture to different data model solutions. However, this solution may add some delays in data retrieval. In the implementation of the dashboard for the two cities, we experienced high waiting times in the rendering phase of some views. Some problems have been overcome with the use of materialized views and others are still being analyzed to obtain more and more effective responses.

In one year of TTD life, we experienced that it is not only a mere collection of visualizations, but it has a strong impact on the city's actors in their daily life. Providing a traffic dashboard to public authorities can guide them through effective analyzes of traffic trends, and gave them insights on how to reduce traffic issues. Moreover, opening the dashboard's visualizations to citizens may encourage their participation and increase transparency in government.

8. Conclusion and future work

The main contribution of the presented work comprises a visual analytics dashboard that provides an efficient means for analyzing urban traffic data in space and time. Additionally, the dashboard enables the exploration of air quality impact among different traffic scenarios by appropriate visual means. This can provide useful insight into the traffic congestion of a particular area in a particular time period and on the impact of the current or new vehicle fleet on urban air quality. Moreover, this paper contributes by providing a flexible framework based on open-source software that can be adaptable in different scenarios, such as other cities but also regions or areas.

This dashboard has been implemented and used in two cities for almost one year, allowing public authorities to grasp important insights from the current traffic scenario and to start defining necessary measures (such as defining low emission zones and/or stimulate electric and hybrid vehicle adoption while gradual phase-out of internal combustion engines) to improve urban air quality. On March 29^{th} 2021, a dissemination event is planned for exposing and further advertising this dashboard to citizens and stakeholders and to gather their feedback.

Future work will explore and compare the use of NoSQL or Graph database to improve efficiency in data storage and retrieval since data retrieval is the core of information visualization. Additional representation and visualization types will be also explored. Among them, hierarchical visualization [37,38] seems promising because it permits to represent the process at different levels of detail. Hierarchical visualizations have great potential for exploration and online monitoring of high-dimensional dynamic data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research reported in this paper was partially supported by the Trafair project 2017-EU-IA-0167, co-financed by the Connecting Europe Facility of the European Union and by the FARD 2021 department project titled "Deep Learning for Urban Event Extraction from News and Social media streams", co-financed by "Enzo Ferrari" Engineering Department of the University of Modena and Reggio Emilia.

References

[1] M. De Gennaro, E. Paffumi, G. Martini, Big data for supporting low-carbon road transport policies in Europe: applications, challenges and opportunities, Big Data Res. 6 (2016) 11–25.
[2] J. Zhang, F. Wang, K. Wang, W. Lin, X. Xu, C. Chen, Data-driven intelligent transportation systems: a survey, IEEE Trans. Intell. Transp. Syst. 12 (2011) 1624–1639.
[3] Air pollution: how it affects our health, https://www.eea.europa.eu/themes/air/health-impacts-of-air-pollution, 2021. (Accessed 15 March 2021).
[4] S.E. Bibri, The IoT for smart sustainable cities of the future: an analytical framework for sensor-based big data applications for environmental sustainability, Sustain. Cities Soc. 38 (2018) 230–253.
[5] Transforming our world: the 2030 agenda for sustainable development, https://www.refworld.org/docid/57b6e3e44.html, 2015. (Accessed 15 March 2021).
[6] T. Fu, A review on time series data mining, Eng. Appl. Artif. Intell. 24 (2011) 164–181.
[7] S. Shekhar, M. Evans, J. Kang, P. Mohan, Identifying patterns in spatial information: a survey of methods, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 1 (2011) 193–214.
[8] A. Gogolou, T. Tsandilas, T. Palpanas, A. Bezerianos, Comparing similarity perception in time series visualizations, IEEE Trans. Vis. Comput. Graph. 25 (2019) 523–533.
[9] G. Atluri, A. Karpatne, V. Kumar, Spatio-temporal data mining: a survey of problems and methods, ACM Comput. Surv. 51 (2018) 83.
[10] N. Andrienko, G. Andrienko, P. Gatalsky, Exploratory spatio-temporal visualization: an analytical review, J. Vis. Lang. Comput. 14 (2003) 503–541, Visual Data Mining.
[11] B. Degraeuwe, P. Thunis, A. Clappier, M. Weiss, W. Lefebvre, S. Janssen, S. Vranckx, Impact of passenger car NOx emissions on urban NO2 pollution – scenario analysis for 8 European cities, Atmos. Environ. 171 (2017) 330–337.
[12] E.-S.M. Zahran, M.J. Smith, L.D. Bennett, 3D visualization of traffic-induced air pollution impacts of urban transport schemes, J. Comput. Civ. Eng. 5 (2013) 452–465, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000198.
[13] N. Tempelmeier, A. Sander, U. Feuerhake, M. Löhdefink, E. Demidova, TA-Dash: an interactive dashboard for spatial-temporal traffic analytics, in: C. Lu, F. Wang, G. Trajcevski, Y. Huang, S.D. Newsam, L. Xiong (Eds.), SIGSPATIAL '20: 28th International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, November 3-6, 2020, ACM, 2020, pp. 409–412, https://doi.org/10.1145/3397536.3422344.
[14] C.V. Gheluwe, I. Semanjski, S. Hendrikse, S. Gautama, Geospatial dashboards for intelligent multimodal traffic management, in: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2020, Austin, TX, USA, March 23–27, 2020, IEEE, 2020, pp. 1–3, https://doi.org/10.1109/PerComWorkshops48775.2020.9156231.
[15] J. Pereira, S. Sargento, J.M. Fernandes, Decision support dashboard for traffic and environment analysis of a smart city, in: M. Helfert, O. Gusikhin (Eds.), Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS 2018, Funchal, Madeira, Portugal, March 16-18, 2018, SciTePress, 2018, pp. 387–394, https://doi.org/10.5220/0006707603870394.
[16] M. Feng, J. Zheng, J. Ren, Y. Liu, Towards big data analytics and mining for UK traffic accident analysis, visualization & prediction, in: ICMLC 2020: 2020 12th International Conference on Machine Learning and Computing, Shenzhen, China, February 15–17, 2020, ACM, 2020, pp. 225–229.
[17] I. Kalamaras, A. Zamichos, A. Salamanis, A. Drosou, D.D. Kehagias, G. Margaritis, S. Papadopoulos, D. Tzovaras, An interactive visual analytics platform for smart intelligent transportation systems management, IEEE Trans. Intell. Transp. Syst. 19 (2018) 487–496.
[18] N. Tempelmeier, Y. Rietz, I.V. Lishchuk, T. Kruegel, O. Mumm, V.M. Carlow, S. Dietze, E. Demidova, Data4UrbanMobility: towards holistic data analytics for mobility applications in urban regions, in: S. Amer-Yahia, M. Mahdian, A. Goel, G. Houben, K. Lerman, J.J. McAuley, R. Baeza-Yates, L. Zia (Eds.), Companion of the 2019 World Wide Web Conference, WWW 2019 San Francisco, CA, USA, May 13-17, 2019, ACM, 2019, pp. 137–145, https://doi.org/10.1145/3308560.3317055.
[19] C. Bachechi, F. Desimoni, L. Po, D.M. Casas, Visual analytics for spatio-temporal air quality data, in: E. Banissi, F. Khosrow-shahi, A. Ursyn, M.W.McK. Bannatyne, N. Datia, K. Nazemi, B. Kovalerchuk, J. Counsell, A. Agapiou, Z. Vrcelj, H. Chau, M. Li, G. Nagy, R. Laing, R. Francese, M. Sarfraz, F. Bouali, G. Venturini, M. Trutschl, U. Cvek, H. Müller, M. Nakayama, M. Temperini, T. Di Mascio, F. Sciarrone, V. Rossano, R. Dörner, L. Caruccio, A. Vitiello, W. Huang, M. Risi, U. Erra, R. Andonie, M. Ahmad, A. Figueiras, A. Cuzzocrea, M. Mabakane (Eds.), 24th International Conference on Information Visualisation, IV 2020, Melbourne, Australia, September 7–11, 2020, IEEE, 2020, pp. 460–466, https://doi.org/10.1109/IV51561.2020.00080.
[20] B. Sudharsan, F. Rollo, L. Po, J.G. Breslin, Air quality sensor network data acquisition, cleaning, visualization, and analytics: a real-world IoT use case, in: UbiComp/ISWC '21: 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2021 ACM International Symposium on Wearable Computers, Virtual Event, September 21–26, 2021, ACM, 2021.
[21] C. Bachechi, L. Po, F. Desimoni, Real-time visual analytics for air quality, in: B. Kovalerchuk, K. Nazemi, R. Andonie, N. Datia, E. Bannissi (Eds.), Integrating Artificial Intelligence and Visualization for Visual Knowledge Discovery, Springer, 2022, https://link.springer.com/book/9783030931186, in press.
[22] P. Grzesik, D. Mrozek, Comparative analysis of time series databases in the context of edge computing for low power sensor networks, in: 20th International Conference, Amsterdam, the Netherlands, June 3-5, 2020, in: Lecture Notes in Computer Science, vol. 12141, Springer, 2020, pp. 371–383.
[23] I. Petre, R. Boncea, C. Radulescu, A. Zamfiroiu, I. Sandu, A time-series database analysis based on a multi-attribute maturity model, Stud. Inform. Control 28 (2019) 177–188.
[24] C. Bachechi, F. Rollo, L. Po, Real-time data cleaning in traffic sensor networks, in: 17th IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2020, Antalya, Turkey, November 2-5, 2020, IEEE, 2020, pp. 1–8, https://doi.org/10.1109/AICCSA50499.2020.9316534.
[25] C. Bachechi, F. Rollo, L. Po, Detection and classification of sensor anomalies for simulating urban traffic scenarios, Cluster Computing (2021), https://doi.org/10.1007/s10586-021-03445-7, in press.
[26] P.Á. López, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. WieBner, Microscopic traffic simulation using SUMO, in: 21st International Conference on Intelligent Transportation Systems, ITSC 2018, Maui, HI, USA, November 4-7, 2018, IEEE, 2018, pp. 2575–2582.
[27] C. Bachechi, L. Po, Implementing an urban dynamic traffic model, in: IEEE/WIC/ACM International Conference on Web Intelligence, WI'19, Oct. 14-17, 2019, ACM, Thessaloniki, Greece, 2019.
[28] L. Po, F. Rollo, C. Bachechi, A. Corni, From sensors data to urban traffic flow analysis, in: 2019 IEEE International Smart Cities Conference, ISC2 2019, Casablanca, Morocco, October 14-17, 2019, IEEE, 2019, pp. 478–485.
[29] D. Gunopulos, G. Das, G. Das, Time series similarity measures and time series indexing (abstract only), SIGMOD Rec. 30 (2001) 624.
[30] C. Bachechi, L. Po, Traffic analysis in a smart city, in: Web4City, International IEEE/WIC/ACM Smart City Workshop: Web for Smart Cities, Thessaloniki, Greece, Oct. 14-17, 2019, 2019.
[31] W. Balzano, A. Murano, L. Sorrentino, S. Stranieri, Network signal comparison through waves parameters: a local-alignment-based approach, in: 5th IEEE International Symposium on Measurements & Networking, M&N 2019, Catania, Italy, July 8–10, 2019, IEEE, 2019, pp. 1–6, https://doi.org/10.1109/IWMN.2019.8805047.
[32] S. Ibarra-Espinosa, R. Ynoue, S. O'Sullivan, E. Pebesma, M.d.F. Andrade, M. Osses, VEIN v0. 2.2: an R package for bottom–up vehicular emissions inventories, Geosci. Model Dev. 11 (2018) 2209–2229.
[33] G. Veratti, S. Fabbi, A. Bigi, A. Lupascu, G. Tinarelli, S. Teggi, G. Brusasca, T.M. Butler, G. Ghermandi, Towards the coupling of a chemical transport model with a micro-scale Lagrangian modelling system for evaluation of urban NOx levels in a European hotspot, Atmos. Environ. 223 (2020) 117285, https://doi.org/10.1016/j.atmosenv.2020.117285, https://www.sciencedirect.com/science/article/pii/S1352231020300273.
[34] L. Ntziachristos, Z. Samaras, C. Kouridis, D. Hassel, I. Mccare, J. Hickman, EMEP/EEA Emission Inventory Guidebook 2009, European Environment Agency (EEA), 2009.
[35] A. Bigi, G. Veratti, S. Fabbi, L. Po, G. Ghermandi, Forecast of the impact by local emissions at an urban micro scale by the combination of Lagrangian modelling and low cost sensing technology: the Trafair project, in: 19th International Conference on Harmonisation Within Atmospheric Dispersion Modelling for Regulatory Purposes, Harmo 2019, 2019.
[36] M. Fu, J. Kelly, J.P. Clinch, Estimating annual average daily traffic and transport emissions for a national road network: a bottom-up methodology for both nationally-aggregated and spatially-disaggregated results, J. Transp. Geogr. 58 (2017).
[37] O. Zoeter, T. Heskes, Hierarchical visualization of time-series data using switching linear dynamical systems, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1202–1214.
[38] N. Li, Z. Jiang, Z. Liu, X. Meng, A method of hierarchical time-series data visualization, in: Proceedings of the 6th International Symposium on Visual Information Communication and Interaction, VINCI '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 113–114.